

MarUja: Prototipo de Asistente Virtual para la Carta de Servicios del Servicio de Informática de la Universidad de Jaén

MarUja: Virtual Assistant prototype for the Catalogue of the Computing Service of the University of Jaén

Eugenio Martínez Cámara L. Alfonso Ureña López José M. Perea Ortega
Departamento de Informática, Escuela Politécnica Superior de Jaén
Universidad de Jaén, E-23071 – Jaén
{emcamara,laurena,jmperea}@ujaen.es

Resumen: Cada vez es mayor la información y el número de servicios que las webs empresariales y de administraciones públicas ofrecen a sus usuarios. Esto hace que la navegación y el acceso a la información sea cada vez más complejo para los usuarios, por lo que se hace necesario facilitar dichas tareas. En este artículo presentamos un prototipo de Asistente Virtual, que es la aplicación práctica de una metodología que permite el desarrollo e implantación de este tipo de sistemas a un coste mínimo.

Palabras clave: Asistentes Virtuales, motor de búsqueda, sintetizador de voz.

Abstract: The information and web services that many organizations offer through its web pages are increasing every day. This makes that navigation and access to information becomes increasingly complex for visitors of these web pages, so, it is necessary to facilitate these tasks for users. In this paper we present a prototype of a Virtual Assistant, which is the result of applying a methodology to develop and set this kind of systems with a minimum cost.

Keywords: Virtual Assistant, search engine, text to speech.

1 Introducción

Un asistente virtual (AV) es una entidad virtual, generalmente representada por un avatar, que es capaz de responder de forma concreta a las consultas expresadas utilizando lenguaje natural por los visitantes de un determinado sitio web.

Los AVs usan técnicas de reconocimiento del lenguaje, lo que les permite interpretar el sentido de las preguntas, seleccionar la respuesta más adecuada, sugerir otras respuestas y conducir al usuario a la web específica en que se encuentra la información que busca.

De forma general los AVs se caracterizan por:

1. Son diseñados para comprender el lenguaje natural.
2. Pueden programarse para conversar en distintos idiomas.
3. Son muy útiles para obtener información sobre los usuarios.
4. Pueden ser diseñados para comprender el contexto de la conversación.
5. Son capaces de generar la respuesta adecuada en función de la consulta.

6. Pueden generar distintos formatos de respuesta, dependiendo del tipo de usuario, y de la pregunta realizada.

Se podría decir que los AV son una nueva generación de motores de búsqueda, que en función de los requerimientos del problema pueden ser más o menos complejos, es decir, en su construcción se pueden utilizar más o menos componentes de PLN. Existen dos grandes grupos de AV:

1. AV con alto grado de PLN: Aquellos en los que se suelen utilizar motores de búsqueda semánticos, gestores de diálogo, perfiles de usuario, u otros componentes de PLN.
2. AV con un bajo grado de PLN: Aquellos en los que se utiliza un reducido número de componentes de PLN. Los más destacados son los que se pueden llamar FAQ's Inteligentes. Estos son un sistema de recuperación de información sobre una base de datos de preguntas y respuestas.

A pesar de sus diferencias internas ambos tienen dos características comunes:

1. Disponen de una animación de aspecto humanoide.
2. Cumplen, de una forma más o menos óptima, el objetivo de facilitar el acceso a la información con una interacción lo más natural posible.

2 Motivación

El Servicio de Informática de la Universidad de Jaén proporciona una serie de servicios a la comunidad universitaria. Para solventar las dudas de los usuarios, en la web del Servicio hay disponible una FAQ¹. El problema reside en el desconocimiento de la comunidad universitaria de la existencia de dicha FAQ, su complicado acceso, y la dificultad de los usuarios para encontrar la respuesta a sus problemas. Por ello la motivación del proyecto se encuentra en facilitar a los usuarios a resolver sus dudas sobre los servicios tecnológicos de la Universidad de Jaén. Por ello, se ha construido un prototipo de AV de bajo grado de PLN para el Servicio de Informática de la Universidad de Jaén.

3 Componentes

Uno de los objetivos del proyecto fue el de minimizar el desarrollo propio, es decir, el de poder construir el sistema a partir de la combinación de componentes software de PLN.

El sistema está formado principalmente por dos componentes software:

1. Motor de búsqueda sintáctico: Se ha utilizado el servidor de búsqueda de código abierto Apache Solr² (Smiley y Pugh, 2009).
2. Sintetizador de voz, Text to Speech Synthesis (TTS): Se seleccionó el motor de TTS del proyecto Festival Speech Synthesis³. Festival no proporciona una voz en español, se tuvo que elegir la voz femenina del proyecto de software libre Hispavoces⁴

4 Arquitectura

Otro de los objetivos fue el de desarrollar un producto con una arquitectura simple que pudiera ser adaptado rápidamente a cualquier

dominio, minimizando al máximo los costes de implantación.

La construcción del sistema comenzó con la extracción automática de las preguntas y respuestas de la FAQ, generando un fichero XML por cada pregunta y respuesta, que contiene la pregunta, la respuesta, y las categorías a las que pertenece la cuestión. Los ficheros XML se utilizan para generar el índice de búsqueda de Solr.

El módulo principal del sistema está escrito en Java, y es el encargado de gestionar la comunicación entre Solr y Festival, por lo que sus funciones son:

1. Recoger las dudas de los usuarios.
2. Preparar y lanzar las preguntas a Solr.
3. Lanzar el resultado más relevante de Solr al servidor TTS, que genera un fichero de audio con la respuesta.
4. Construir la respuesta del sistema, con el audio generado por Festival, y tres preguntas similares a la realizada.

5 Conclusiones

En este trabajo se ha descrito el prototipo MarUja, un AV desarrollado para el Servicio de Informática de la Universidad de Jaén. El principal objetivo fue el de proporcionar una interfaz en lenguaje natural para satisfacer las necesidades de información de los usuarios relacionadas con las prestaciones de dicho departamento. Además, presentamos la metodología seguida para generar el prototipo.



Figura 1: AV MarUja

Agradecimientos

Investigación financiada por el Fondo Europeo de Desarrollo Regional (FEDER), proyecto TEXT-COOL 2.0 (TIN2009-13391-C04-02) del Gobierno Español.

Bibliografía

Smiley, D. y Pugh, E. 2009. Solr 1.4 Enterprise Search Server. Packt Publishing Ltd.

¹ <http://faq.ujaen.es/>

² <http://lucene.apache.org/solr/>

³ <http://www.cstr.ed.ac.uk/projects/festival/>

⁴ <http://forja.guadalinex.org/projects/hispavoces/>